

METHODS FOR CREATING A COMPOUND LIBRARY

5

Cross-reference to Related Applications

The present application is a continuation-in-part of U.S. Application Serial No. 09/677,107 filed on September 29, 2000, which claims the benefit of Serial Nos. 60/156,818, filed on September 29, 1999, 60/161,682, filed on October 26, 1999, and 60/192,685, filed on March 28, 2000, each of which is 10 incorporated herein by reference in its entirety.

Background of the Invention

From an organic chemistry standpoint, the process of drug design can be considered to involve two steps. First, a lead chemical template (often one or more) is selected. Second, a synthetic chemistry effort is undertaken to create 15 analogs of the lead chemical template to create a compound or compounds possessing the desired therapeutic and pharmacokinetic properties.

An important step in the drug discovery process is the selection of a suitable lead chemical template upon which to base a chemistry analog program. The process of identifying a lead chemical template for a given molecular target 20 typically involves screening a large number of compounds (often more than 100,000) in a functional assay, selecting a subset based on some arbitrary activity threshold for testing in a secondary assay to confirm activity, and then assessing the remaining active compounds for suitability of chemical elaboration.

This process can be quite time- and resource-consuming, and has 25 numerous disadvantages. It requires the development and implementation of a high-throughput functional assay, which by definition requires that the function of the molecular target be known. It requires the testing of large numbers of compounds, the vast majority of which will be inactive for a given molecular target. It leads to the depletion of chemical resources and requires the continual 30 maintenance of large collections of compounds. Importantly, it often leads to a final pool of potential lead templates that for the most part, with the exception of affinity for a given molecular target, do not possess desirable drug-like qualities. In some cases, high-throughput functional assays do not identify any compounds

DOCKET NUMBER

from the large number (e.g., 100,000) of compounds screened that meet the criteria established for activity.

Thus, what is needed is a faster and better approach to identifying a lead chemical template.

5

Summary of the Invention

The present invention is related to rational drug design. Specifically, the present invention provides an approach to the development of a library of compounds as well as methods for identifying compounds (e.g., ligands) that bind to a specific target molecule (e.g., proteins) and lead chemical templates that can be used, for example, in drug discovery and design. Significantly and preferably, this approach for identifying ligands for target molecules (e.g., proteins) uses nuclear magnetic resonance (NMR) spectroscopy. There are numerous NMR spectroscopic techniques currently available that detect binding of small molecules to targets such as protein targets, including targets identified using genomics techniques that lack a functional assay. Ligands with only moderate binding affinities, which might be overlooked in a traditional functional assay but yet might serve as templates for subsequent synthetic chemistry efforts, can potentially be identified using the present invention.

10

15

20

25

30

In one aspect, the present invention provides a method of creating a chemical compound library, and the library itself. The method includes:

selecting compounds having a molecular weight of no greater than about 350 grams/mole; and selecting compounds having a solubility in deuterated water of at least about 1 mM at room temperature. Preferably, a majority (i.e., greater than 50%) of the compounds in the chemical compound library have a molecular weight of no greater than about 350 grams/mole and a solubility in deuterated water of at least about 1 mM at room temperature. More preferably, at least about 75% of the compounds, and most preferably, all of the compounds in the chemical compound library have a molecular weight of no greater than about 350 grams/mole and a solubility in deuterated water of at least about 1 mM at room

temperature. Preferably, this library of compounds includes at least about 75 compounds, more preferably, at least about 300 compounds, and most preferably, at least about 2000 compounds, and have relatively diverse chemical structures. Herein, the molecular weights of the compounds are determined 5 without solubilizing counterions (if the compounds are salts) and without water molecules of hydration. Also, concentrations are reported based on aqueous solutions, which may or may not include a buffer.

In another embodiment, the present invention provides a method of identifying a lead chemical template (of which there often may be one or more), 10 for example, for designing a bioactive agent such as a drug (e.g., a compound having therapeutic and/or prophylactic capabilities). The method includes: selecting compounds having a molecular weight of no greater than about 350 grams/mole, and a solubility in deuterated water of at least about 1 mM at room temperature to create a chemical compound library; identifying at least one 15 compound from the library that functions as a ligand (i.e., a compound that binds to a target molecule) having a dissociation constant to a target molecule (e.g., protein) of no weaker than (i.e., at least) about 100 μM ; and using the ligand to identify a lead chemical template, which can be used, for example, for designing a drug. Preferably, the lead chemical template has a dissociation constant to a 20 target molecule (e.g., protein) of no weaker than (i.e., at least) about 1 μM . Preferably, the lead chemical template can be identified through further screening efforts or through direct chemical elaborations. Preferably, a majority (i.e., greater than 50%) of the compounds in the chemical compound library, more preferably, at least about 75%, and most preferably, all of the compounds 25 in the chemical compound library, have a molecular weight of no greater than about 350 grams/mole and a solubility in deuterated water of at least about 1 mM at room temperature.

Another embodiment of the present invention provides a method of identifying a compound that binds to a target molecule (e.g., protein). The 30 method includes: providing a plurality of mixtures of test compounds, each mixture being in a (separate) sample reservoir (preferably, a sample reservoir of a multiwell sample holder (e.g., a 96-well microtiter plate)); introducing a target

TOP SECRET - ESTATE PROTECTOR

molecule (e.g., protein) into each of the sample reservoirs to provide a plurality of test samples; providing a nuclear magnetic spectrometer equipped with a flow-injection probe; transferring each test sample from the sample reservoir into the flow-injection probe; collecting a relaxation-edited (preferably, a one-dimensional (1D) relaxation-edited) nuclear magnetic resonance spectrum (preferably, a ^1H NMR spectrum) on each sample in each reservoir; and comparing the spectra of each sample to the spectra taken under the same conditions in the absence of the target molecule (e.g., protein) to identify compounds that bind to the target molecule (e.g., protein); wherein the

5 concentration of target molecule (e.g., protein) and each compound in each sample is no greater than about 100 μM . Preferably, the mixture of compounds comprises at least about 3 compounds (more preferably, at least about 6 compounds, and most preferably, at least about 10 compounds), each having at least one distinguishable resonance in an NMR spectrum (preferably, a 1D NMR

10 spectrum, and more preferably, a 1D ^1H NMR spectrum) of the mixture.

15

Preferably, in this method, the ratio of target molecule (e.g., protein) to compounds in each sample reservoir is about 1:1. More preferably, the concentration of target molecule (e.g., protein) and each compound in each sample is at least about 25 μM . Most preferably, the concentration of target

20 molecule (e.g., protein) and each compound in each sample is no greater than about 50 μM .

Sample requirements can be reduced even further if WaterLOGSY (water-ligand observation with gradient spectroscopy) methods are used as an alternative to the relaxation-editing method described above to detect the binding

25 interaction.

The present invention provides yet another method of identifying a compound that binds to a target molecule (e.g., protein). This method includes: providing a plurality of mixtures of test compounds, each mixture being in a sample reservoir; introducing a target molecule into each of the sample

30 reservoirs to provide a plurality of test samples; providing a nuclear magnetic resonance spectrometer equipped with a flow-injection probe; transferring each test sample from the sample reservoir into the flow-injection probe; collecting a

WaterLOGSY nuclear magnetic resonance spectrum (preferably, a 1D WaterLOGSY nuclear magnetic resonance spectrum) on each sample in each reservoir; and analyzing the spectra of each sample to distinguish binding compounds from nonbinding compounds by virtue of the opposite sign of their water-ligand nuclear Overhauser effects (NOEs). Preferably, the concentration of each compound in each sample is no greater than about 100 µM, although higher concentrations can be used if desired.

In this method when binding is detected using the WaterLOGSY technique, extremely low levels of target can be used with ratios of ligand to target of about 100:1 to about 10:1. Preferably, the concentration of target molecule is no greater than about 10 µM. More preferably, the concentration of target molecule is about 1 µM to about 10 µM. For data analysis, binding compounds are distinguished from nonbinders (i.e., nonbinding compounds) by the opposite sign of their water-ligand NOEs. With this method, there is no need to collect a reference spectrum in the absence of a target molecule.

In preferred embodiments of the present invention, a majority of the compounds in the library have a solubility in deuterated water of at least about 1 mM at room temperature (i.e., about 25°C to about 30°C), and a molecular weight of no greater than about 350 grams/mole. For effective use of a compound identified as a ligand for a given target in the search for a lead chemical template, preferably, the dissociation constant of the identified ligand to a target molecule is no weaker than (i.e., at least) about 100 µM. For effective use of a lead chemical template in further drug design, preferably, the dissociation constant for the lead chemical template to a target molecule is no weaker than (i.e., at least) about 1 µM.

In another aspect, the invention provides a method of identifying a protein function. The method includes providing a plurality of mixtures of test compounds, each mixture being in a sample reservoir and containing a plurality of test compounds; introducing a target molecule into each of the sample reservoirs to provide a plurality of test samples; providing a nuclear magnetic resonance spectrometer equipped with a flow-injection probe; transferring each test sample from the sample reservoir into the flow-injection probe; collecting a

TOP SECRET//COMINT

relaxation-edited nuclear magnetic resonance spectrum on each sample in each reservoir; comparing the spectra of each sample to the spectra taken under the same conditions in the absence of the target molecule to identify compounds that bind to the target molecule, wherein the concentration of target molecule and

5 each compound in each sample is no greater than about 100 μM ; and determining a function of the target molecule based upon known binding characteristics of the test compounds that bind to the target molecule.

Brief Description of the Drawings

10 Figure 1. Schematic diagram illustrating the use of NMR to discover a ligand having an approximate dissociation constant of $1.0 \times 10^{-4} \text{ M}$ (left figure), to use the discovered ligand to direct the discovery of a lead chemical template having an approximate dissociation constant of $1.0 \times 10^{-6} \text{ M}$ (middle figure), and then via synthetic chemistry and structure-directed drug design arrive at a drug

15 candidate having an approximate dissociation constant of $1.0 \times 10^{-8} \text{ M}$.

Figure 2. Comparison of the two-dimensional HA (hydrogen-bond acceptor) vs. CHRG (charge) BCUT plots for the compounds contained in the NMR library described herein (dark squares) and a larger chemical library database (gray spots).

20 Figure 3A. One-dimensional relaxation-edited ^1H NMR spectrum of a compound set containing three compounds designated (1), (2), and (3). Resonances are numbered corresponding to the individual components in the set.

Figure 3B. One-dimensional relaxation-edited ^1H NMR spectrum of the same set of compounds shown in Figure 3A in the presence of flavodoxin.

25 Arrows identify resonances that experience a significant reduction in intensity.

Figure 4A. Region of the 2D ^1H - ^{15}N HSQC spectrum of flavodoxin alone and in the presence of a 10-fold excess of compound (1). Residues with significant chemical shift changes in the presence of (1) are boxed and labeled with their amino acid type and sequence number.

30 Figure 4B. Secondary structure representation of the flavodoxin global fold. The flavin cofactor is shown in stick format. Residues with the largest chemical shift changes in the presence of (1) are shown in white.

Figure 5A. One-dimensional relaxation-edited ^1H NMR spectrum of a compound set containing three compounds in the presence of flavodoxin.

Figure 5B. One-dimensional relaxation-edited ^1H NMR spectrum of the same compound set shown in Figure 5A in the presence of the antibacterial target protein. Arrows identify resonances from Ligand A (Figure 6) that experience a significant reduction in intensity in the presence of the antibacterial target protein.

Figure 6. IC_{50} values of the original ligand, Ligand A, and four structurally related compounds, Ligands B-E, identified in a similarity search based on the structure of Ligand A.

Figure 7. Region of the 2D ^1H - ^{15}N HSQC spectrum of the antibacterial target protein alone and in the presence of a 10-fold excess of Ligand A. Several resonances with large chemical shift changes in the presence of Ligand A are boxed and labeled with their amino acid sequence number.

Figure 8A. One-dimensional relaxation-edited ^1H NMR spectrum of a compound set containing ten compounds.

Figure 8B. One-dimensional relaxation-edited ^1H NMR spectrum of the same set of compounds in Figure 8A in the presence of the antiviral target protein. Arrows identify resonances, all belonging to the same compound, that experience a significant reduction in intensity in the presence of the antiviral target protein.

Figure 9. Region of the 2D ^1H - ^{15}N HSQC spectrum of the antiviral target protein alone and in the presence of the ligand identified from Figure 8. Several resonances with large chemical shift changes in the presence of this ligand are boxed and labeled with their amino acid sequence number.

Figure 10. Schematic of the BEST flow system: (1) computer workstation, (2) NMR console, (3) Gilson sample handler, (4) flow probe in the magnet, and (5) nitrogen gas. The Gilson sample handler is labeled as follows: (A) keypad, (B) syringe, (C) injector, (D) solvent reservoir, (E) solvent rack, (F) sample racks, (G) waste reservoir, (H) Rheodyne valves, (I) injection port, and (J) recovery unit.

Figure 11. Schematic of a Bruker flow probe showing (A) the total probe volume, (B) the flow cell volume, and (C) the positioning volume.

Figure 12. 600.13 MHz ^1H NMR spectra of a 100 μM NMR library sample with the positioning volume set to (A) -100 μl , (B) 0 μl , and (C) +100 μl .

Figure 13. Overlay of the two-dimensional HA (hydrogen-bond acceptor) vs. CHRG (charge) BCUT plots for the compounds in the CMC index (gray) and the lead-like compounds contained therein (black).

Figure 14. Regions of the 600.13 MHz relaxation-edited ^1H NMR spectra of a nine compound mixture (A) without and (B) with added target protein. Protein and each ligand were 50 μM . Spectra were acquired on a Bruker 5 mm flow-injection probe at 27°C. A total of 1K scans were collected resulting in a total acquisition time of about 60 minutes per spectrum. A relaxation filter of 174 milliseconds (ms) was used. Arrows identify resonances that disappear in the presence of protein.

Figure 15. Regions of the 600.13 MHz relaxation-edited ^1H NMR spectra of a single compound (A) without and (B) with added target protein. Protein and ligand were 50 μM . Spectra were acquired on a regular Bruker 5 mm TXI probe at 27°C. A total of 512 scans were collected resulting in a total acquisition time of about 30 minutes per spectrum. A relaxation filter of 174 ms was used.

Figure 16. Region of the 600.13 MHz WaterLOGSY spectrum of a compound mixture with added target protein. The concentration of protein was 10 μM while the concentration of each compound was 100 μM . The spectrum was acquired on a Bruker 5 mm flow-injection probe at 27°C. A total of 4K scans were collected resulting in a total acquisition time of about 288 minutes.

A mixing time of 2.0 seconds was used.

Figure 17. Comparison of WaterLOGSY spectrum (bottom panel) of thrombin with a compound mixture of the genomics screening library and the reference spectrum of DPS (top panel).

Jnd C1
30

Detailed Description of Preferred Embodiments of the Invention

The present invention involves the selection of a generally small library of structurally diverse compounds that are generally water soluble, have a

relatively low molecular weight, and are amenable to synthetic chemistry elaboration. Significantly and advantageously, for certain embodiments, the present invention preferably involves carrying out a binding assay at relatively low concentrations of target and near equimolar ratios of ligand to target, or even 5 at extremely low concentrations of target and higher ratios of ligand to target.

In a method of the present invention, a relatively small subset of compounds (preferably, at least about 75, more preferably, at least about 300, most preferably, at least about 2000, and typically no more than about 10,000) that mimics the structural diversity of compounds in much larger collections is 10 created based on a predetermined set of criteria. This generally small library is screened for binding affinity to a target molecule (as determined herein by dissociation constants). The compounds from the library that are identified to be effective ligands (typically, having an affinity for a desired target as evidenced by a dissociation constant of at least about 1.0×10^{-4} M) are then used to focus 15 further screening efforts or to direct chemical elaborations to arrive at one or more lead chemical templates (which, typically have an affinity for a desired target as evidenced by a dissociation constant of at least about 1.0×10^{-6} M).

This process is shown schematically in Figure 1.

Significantly, time and resources are saved by screening far fewer 20 compounds using the present invention. Use of a binding assay, such as the one based on NMR spectroscopy described herein, eliminates the need to develop a high-throughput functional assay, and also allows the methods to be used on molecular targets lacking a known function.

Thus, the present invention provides methods of identifying a compound 25 that binds to a target molecule (preferably, a protein) that are based on NMR spectroscopy techniques. Such methods typically involve the use of relaxation-editing techniques, for example, which involve monitoring changes in resonance intensities (preferably, significant reductions in intensities) of the test compound upon the addition of a target molecule. Preferably, the relaxation-editing 30 techniques are one-dimensional, and more preferably, one-dimensional ^1H NMR techniques. Alternatively, such methods can involve the use of WaterLOGSY. This involves the transfer of magnetization from bulk water to detect the binding interaction. Using WaterLOGSY techniques, binding compounds are

distinguished from nonbinders by the opposite sign of their water-ligand nuclear Overhauser effects (NOEs).

Important elements that contribute to the success of the methods of the invention preferably include developing a suitable small library of compounds to screen, carrying out the binding assay at low concentrations of target and near equimolar ratios of ligand to target (for relaxation-editing), or at extremely low concentrations of target (if desired) and higher ratios of ligand to target (for WaterLOGSY), and the capacity for rapid throughput of data collection. For example, for relaxation-editing NMR techniques, the concentration of target molecule is preferably no greater than about 1.0×10^{-4} M, and for WaterLOGSY NMR techniques, the concentration of target molecule is preferably no greater than about 10 μ M.

The selection of compounds in a small library (preferably, at least about 75 compounds, more preferably, at least about 300 compounds, and most preferably, at least about 2000 compounds) is important in that its diversity should mimic the diversity of larger compound collections. Preferably, each component possesses many of the desirable qualities of a lead chemical template. These include water solubility, low molecular weight (preferably, no greater than about 350 grams/mole, more preferably, no greater than about 325 grams/mole, and most preferably, less than about 325 grams/mole), and amenability to synthetic chemistry elaboration. Templates possessing these qualities, as compared to a template selected randomly, are preferably considered to be predisposed to being lead-like and having an increased likelihood of ultimately leading to a drug.

Good structural diversity in a library increases the likelihood that one or more compounds will possess structural characteristics important for binding to a given molecular target. Predisposing the compounds to be water soluble, to have low molecular weight (preferably, no greater than about 350 grams/mole, more preferably, no greater than about 325 grams/mole, and most preferably, less than about 325 grams/mole), and to be amenable to synthetic elaboration increases the likelihood that a compound found to be a ligand will lead to a related compound or compounds suitable as a lead chemical template for use, for example, in a process of identifying an effective therapeutic and/or prophylactic agent.

Additionally, the requirement for good water solubility (preferably, at least about 1.0 x 10⁻³ M in deuterated water at room temperature) is important in that it increases the likelihood of success of other downstream drug-design projects, such as co-crystallization attempts, calorimetry studies, and enzyme kinetic analyses.

Carrying out a relaxation-editing binding assay (preferably, a 1D ¹H NMR assay) at low concentrations of target (preferably, no greater than about 1.0 x 10⁻⁴ M, and more preferably, no greater than about 5.0 x 10⁻⁵ M) and near equimolar ratios of ligand to target creates the requirement that compounds testing positive for binding have affinities within a factor of about 3-4 of this same concentration (preferably, having a dissociation constant of no less than about 2.0 x 10⁻⁴ M). A similar affinity threshold can be obtained by carrying out a WaterLOGSY based binding assay at even lower target concentrations (preferably, no greater than about 10 μM, but is more preferably about 1 μM to about 10 μM) and ligand to target ratios of about 100:1 to about 10:1. This level of affinity is desired if the subsequent steps of focused screening and directed chemical elaboration are to be successful in elucidating a lead chemical template with very low affinity (e.g., one having a dissociation constant of at least about 1.0 x 10⁻⁶ M). Carrying out the initial screening at these low concentrations also avoids detection of unwanted compounds with much smaller dissociation constants in the 1.0 x 10⁻³ M range, which are less specific in their binding and therefore harder to turn into lead chemical templates given their weak affinity initially.

The capacity for rapid throughput of data collection is important if a large number of molecular targets are to be screened. Preferably, flow NMR techniques can reduce the amount of time and effort required to evaluate small molecules for binding to a given target. For example, the use of a Bruker Efficient Sample Transfer system in combination with a tubeless, flow-injection NMR probe has proven to be much faster and less labor intensive than the use of traditional NMR tubes. A significant increase in throughput is obtained compared to both manual sample changing and to using an autosampler. Implementation of the screening process using multiwell sample holders also

standardizes the experimental setup as well as the components in a given mixture from one molecular target to the next.

The following is a description of a preferred method for carrying out the present invention. It is provided for exemplification purposes only and should
5 not be considered to unnecessarily limit the invention as set forth in the claims.

In the design of a preferred small library of structurally diverse compounds according to the present invention, compounds were selected from a large library based on dissimilarity, predicted water solubility, low molecular weight, and chemical intuition. Some were based on frameworks suggested in
10 the literature, although some literature-suggested frameworks were consciously avoided. Each compound was tested for solubility at 1.0×10^{-3} M in $^2\text{H}_2\text{O}$ and for purity by mass spectrometry and ^1H NMR spectroscopy. Compounds deemed to be water soluble and pure were kept for inclusion in the final library (approximately 30% of the initial compounds). The resulting library contains
15 approximately 300 compounds. One measure of the degree of structural diversity of the compounds in this small library is shown in Figure 2. This is based on the technique described in Pearlman et al., *Perspectives in Drug Discovery & Design*, 9, 339-353 (1998). Preferably, the compound library includes compounds of sufficiently diverse chemical structure that one would
20 expect at least one compound to bind to a given target protein with an affinity (dissociation constant) no weaker than (i.e., at least) about 200 μM . Herein, compounds of diverse chemical structure are those that have a variety of backbone hydrocarbon structures (e.g., linear, branched, cyclic – which may or may not be aromatic, have fused rings, etc.), optionally including a variety of
25 heteroatoms (e.g., oxygen, nitrogen) and a variety of functional groups (e.g., carbonyls) in a variety of positions (e.g., pointing in various directions at a variety of distances from each other). Ideally, using the technique described in Pearlman et al., *Perspectives in Drug Discovery & Design*, 9, 339-353 (1998), the library of compounds displays a pattern of well-dispersed black squares (e.g.,
30 see Figure 2).

In order to increase the throughput of the NMR screening, compounds were grouped into 32 sets of 6-10 compounds that have at least one distinguishable resonance in a 1D ^1H NMR spectrum of the mixture. To

accomplish this, a 1D ¹H NMR spectrum was obtained of each mixture in 100% ²H₂O and in 0.1 M sodium phosphate/100% ²H₂O at pH 6.5. Two solvents were used in order to determine the assignment of pH-tritatable resonances in the spectrum. Each of the 32 mixtures was then plated out into separate wells of a 5 96-well plate, using 25 µL of a 1.0 x 10⁻³ M solution, and frozen at -80°C until needed. In an initial version of the NMR screening library, approximately 70 compounds were grouped into 21 sets of 3-4 compounds each.

After a 96-well plate had completely thawed, a solution containing a molecular target protein was added to each well containing a mixture of 10 compounds in the 96-well plate. The final concentration of protein is typically about 5.0 x 10⁻⁵ M. The ratio of each compound in a mixture to protein is typically about 1:1. This process typically involves adding 475 mL of protein to each mixture. Dispersion throughout the mixture was facilitated by shaking the 96-well plate for 20 minutes following addition of protein.

15 A 1D relaxation-edited ¹H NMR spectrum was collected on each protein/compound mixture solution using a Bruker DRX600 or a Bruker AMX400 spectrometer equipped with a shielded magnet, a Gilson sample handler, and a 5 mm (250 µL sample cell) flow-injection NMR probe. The use of a shielded magnet greatly reduces the magnetic fringe field surrounding the 20 high field magnet and allows the Gilson sample handler to be placed in close proximity to the magnet. The Gilson liquid sample handler transfers samples from 96-well plates into the flow-injection probe and, if desired, returns the samples back to the 96-well plate. A compound or compounds that bind to a given target are identified by comparing the 1D relaxation-edited ¹H NMR 25 spectrum collected in the presence of added protein to that of the identical mixture of compounds in the absence of protein. A compound is identified as a ligand for a given target if one or more of its resonances (preferably ¹H resonance or resonances) are significantly reduced (i.e., greater than about 75% reduction in one or more resonances) in intensity in the presence of target 30 molecule (e.g., protein) as compared to the spectrum collected in an identical fashion in the absence of target molecule (e.g., protein).

TOP SECRET - GOVERNMENT USE ONLY

Sample requirements can be reduced even further if WaterLOGSY methods are used as an alternative to the relaxation-editing method described above to detect the binding interaction. WaterLOGSY is described in more detail in C. Dalvit et al., *J. Biomol. NMR*, **18**, 65-68 (2000).

Spec 5

Since the WaterLOGSY experiment relies on the transfer of magnetization from bulk water to detect the binding interaction, it is a very sensitive technique. As such, the concentration of target molecule (e.g., protein)

Spec 3

in each sample preferably can be reduced to no greater than about 10 μM (preferably, about 1 μM to about 10 μM) while the concentration of each

10 compound can be about 100 μM . This results in ratios of target molecule to compounds in each sample reservoir of about 100:1 to about 10:1. The exact concentrations and ratios used can vary depending on the size of the target molecule, the amount of target molecule available, the desired binding affinity detection limit, and the desired speed of data collection. In contrast to the

15 relaxation-editing method, there is no need to collect a comparison or control spectrum to identify binding compounds from nonbinders. Instead, binding compounds are distinguished from nonbinders by the opposite sign of their water-ligand nuclear Overhauser effects (NOEs).

Ligand binding was confirmed by making fresh solutions containing only

20 the identified ligand, with and without added protein at a 1:1 ratio, and comparing the 1D relaxation-edited ^1H NMR spectra. In addition, the ligand's dissociation constant was estimated by analyzing several 1D diffusion-edited ^1H NMR spectra collected at several gradient strengths. The relative diffusion coefficients for the protein, for the ligand in the presence of protein, and for the

25 ligand in the absence of protein, in conjunction with known protein and ligand concentrations, were used to estimate the ligand's dissociation constant. These spectra are typically collected using an NMR spectrometer, a conventional high resolution probe, and regular 5 mm NMR tubes.

Once a ligand had been identified and confirmed, its structure is used to

30 identify available compounds with similar structures to be assayed for activity or affinity, or to direct the synthesis of structurally related compounds to be assayed for activity or affinity. These compounds are then either obtained from inventory

or synthesized. Most often, they are then assayed for activity using enzyme assays. In the case of molecular targets that are not enzymes or that do not have an enzyme assay available, these compounds can be assayed for affinity using NMR techniques similar to those described above, or by other physical methods 5 such as isothermal denaturation calorimetry. Compounds identified in this step with affinities for the molecular target of about 1.0×10^{-6} M are typically considered lead chemical templates.

In some instances, ligand binding is further studied using more complex NMR experiments or other physical methods such as calorimetry or X-ray 10 crystallography. These downstream studies have a greater chance of success since the ligands and lead chemical templates so identified are fairly water soluble. For instance, if [^{15}N]protein is available, 2D ^1H - ^{15}N HSQC (heteronuclear single quantum correlation) spectra can be collected with and 15 without added ligand to locate the ligand's binding site on the protein. In cases where the protein is small enough (molecular weight less than about 30,000) and further characterization of protein/ligand interactions is desired, 3D NMR experiments can be carried out on [^{13}C / ^{15}N]protein/[^{12}C / ^{14}N]ligand complexes. Attempts to soak lead chemical templates identified by this method into existing 20 protein crystals, or to form co-crystals, can also be carried out.

20

Examples

Objects and advantages of this invention are further illustrated by the following examples, but the particular materials and amounts thereof recited in these examples, as well as other conditions and details, should not be construed 25 to unduly limit this invention.

Example 1. Use of NMR Spectroscopy to Identify Ligands for Flavodoxin

Reference 1D ^1H NMR spectra of the individual compounds and combinations of compounds were recorded in $^2\text{H}_2\text{O}$ solution on a Bruker ARX- 30 400 spectrometer. One-dimensional relaxation-edited ^1H NMR spectra of samples containing a mixture of flavodoxin and a given compound combination were recorded in $^2\text{H}_2\text{O}$ solution on a Bruker DRX-500 spectrometer. A spin lock time of 350 milliseconds was used. The screening experiments were carried out

TOCTTFTTTTFTTTT

on solutions that were 5.0×10^{-5} M flavodoxin and 1.0×10^{-4} M of each ligand present. Two-dimensional $^1\text{H}-^{15}\text{N}$ HSQC spectra were recorded in $^1\text{H}_2\text{O}$ solution on a Bruker DRX-500 spectrometer. Samples were 5.0×10^{-5} M flavodoxin with a 3-10 fold excess of a given ligand. All solutions containing flavodoxin were 5 buffered with 1.0×10^{-2} M phosphate at pH 6.4. The *Desulfovibrio vulgaris* flavodoxin used in all experiments was ^{15}N -enriched.

To create the NMR ligand screening library, an initial set of compounds was selected by a search of a larger library of compounds based on dissimilarity, predicted water solubility, low molecular weight (preferably, no greater than 10 about 350 grams/mole, more preferably, no greater than about 325 grams/mole, and most preferably, less than about 325 grams/mole), and chemical intuition. These compounds were then tested for water solubility and purity. Compounds with no visible precipitate or suspension at a concentration of 1.0×10^{-3} M were deemed to be water soluble. Compounds with the predicted parent ion molecular 15 weight and otherwise normal mass spectra were deemed to be pure. Reference 1D ^1H NMR spectra were collected on compounds meeting these criteria. Combinations of three or four compounds were then assembled in which at least one distinguishing ^1H NMR resonance for each compound could be readily identified. A reference 1D ^1H NMR spectrum was then recorded for each 20 combination of compounds. As an example, three compounds, designated here as (1), (2), and (3), were combined into one set. The 1D ^1H NMR spectrum of this combination set is illustrated in Figure 3A. Resonances from each of the individual components are readily identified, especially in the aliphatic region of the spectrum. At the time of this work, the NMR ligand library contained 25 approximately 70 compounds incorporated into 21 unique assortments containing three or four compounds each.

One-dimensional relaxation-edited ^1H NMR spectroscopy was used to screen the library for binding to the model target protein, *Desulfovibrio vulgaris* flavodoxin. For most of the compound combinations in the presence of 30 flavodoxin, there was little or no reduction in resonance intensity with the 350-millisecond spin-lock time. However, for two of the compound combinations, the intensities of resonances corresponding to one of the compounds in the mixture were significantly reduced. Figure 3B exemplifies this for the same

TOP SECRET - 6283.NCP2

combination illustrated in Figure 3A. The resonances corresponding to (2) and (3) are not affected by the spin-lock filter in the presence of flavodoxin. However, the two aliphatic resonances of (1) at 1.8 ppm and 3.7 ppm are significantly reduced in intensity by the spin-lock filter in the presence of 5 flavodoxin, indicating that (1) is binding to the protein. Similar experiments indicated that a second compound, contained within a different combination of compounds, also binds to flavodoxin. These were the only two compounds among those tested that clearly bind to flavodoxin.

Two-dimensional ^1H - ^{15}N HSQC spectra were subsequently recorded on 10 [^{15}N]flavodoxin to further investigate the interaction of these two ligands with the protein. Since amide backbone ^1H and ^{15}N resonance assignments for this protein are known (Stockman et al., *J. Biomol. NMR*, **3**, 133-149 (1993)), analysis of the ligand-induced changes in ^1H and ^{15}N chemical shifts could be used to identify the ligand binding sites. Typical chemical shift changes 15 observed are delineated in Figure 4A, which shows an overlay of the ^1H - ^{15}N HSQC spectra of flavodoxin alone and in the presence of excess (1). Residues with the largest ligand-induced chemical shift changes are indicated in white on the structure of the protein (Watt et al., *J. Mol. Biol.*, **218**, 195-208 (1991)) in Figure 4B. Compound (1) binds near the flavin cofactor binding site. 20 Interestingly, the binding sites as defined by this data for the two ligands identified are at adjacent, partially overlapping locations on the surface near the flavin cofactor binding site.

Example 2. Use of NMR Spectroscopy to Identify a Lead Chemical Template for an Antibacterial Target Protein

Numerous protein targets are amenable to an NMR process of identifying a lead chemical template. In this example, the technique is illustrated for an antibacterial target protein with a molecular weight of about 20 kDa.

All solutions containing the antibacterial target protein were buffered 30 with 2.5×10^{-2} M phosphate at pH 7.4. The protein used for the 1D screening and dissociation constant determination experiments was unlabeled, while that used for the 2D ^1H - ^{15}N HSQC experiments was ^{15}N -enriched.

One-dimensional relaxation-edited ^1H NMR spectra of samples containing a mixture of the target protein and a given compound combination were recorded in $^2\text{H}_2\text{O}$ solution on a Bruker DRX-500 spectrometer. A spin lock time of 350 milliseconds was used. The screening experiments were carried out 5 on solutions that were 1.0×10^{-4} M target protein and 1.0×10^{-4} M of each ligand. The library used for the screening process was identical to that described in Example 1.

Two-dimensional ^1H - ^{15}N HSQC spectra were recorded in $^1\text{H}_2\text{O}$ solution on a Bruker DRX-500 spectrometer. Samples contained 8.0×10^{-5} M target 10 protein with a 9-10 fold excess of a given ligand.

Ligand dissociation constants were estimated by determining relative diffusion coefficients for target protein alone, ligand in the absence of target protein, and ligand in the presence of target protein (Lennon et al., *Biophys. J.*, 67, 2096-2109 (1994)). Relative diffusion coefficients were determined using 15 pulsed-field-gradient NMR experiments incorporating a bipolar longitudinal eddy-current delay sequence (Wu, *J. Magn. Reson. Ser. A*, 115, 260-264 (1995)).

One-dimensional relaxation-edited ^1H NMR spectroscopy was used to screen the small molecule library for binding to this target protein in a manner analogous to that previously described in Example 1. With this technique, a 20 reduction in resonance intensity is observed if a compound interacts with the target protein, thus identifying it as a ligand. For most of the compound combinations in the presence of the antibacterial target protein, there was little or no reduction in resonance intensity with the 350-millisecond spin-lock time. However, for some of the compound combinations, the intensities of resonances 25 corresponding to one of the compounds in the mixture were significantly reduced. The results from one such compound combination are described here.

As a control, the 1D relaxation-edited ^1H NMR spectrum of a certain mixture in the presence of a different protein, flavodoxin, is shown in Figure 5A. All ligand resonances are observed with full intensity. The corresponding 1D 30 relaxation-edited ^1H NMR spectrum of this same mixture acquired in the presence of the antibacterial target protein is shown in Figure 5B. The intensities of all resonances corresponding to Ligand A in Figure 5B are clearly reduced in the presence of the antibacterial target protein. This indicates that Ligand A is

TOP SECRET - FEDERAL BUREAU OF INVESTIGATION

binding to the protein. The binding is specific to the antibacterial target protein since the resonance intensities are not reduced in the presence of flavodoxin.

Binding of Ligand A was confirmed by repeating the relaxation-filtered experiments on a solution containing protein and just Ligand A. Using this same sample, as well as samples of protein alone and Ligand A alone, a separate set of experiments that use pulsed-field-gradient techniques was collected to determine relative diffusion coefficients. From this data, the dissociation constant for Ligand A was estimated by NMR measurements to be approximately 1.4×10^{-4} M.

In order to ascertain whether the binding of Ligand A and structurally related analogs inhibited the activity of this enzyme, and if so to what degree, IC₅₀ values were determined. To determine IC₅₀ values, various concentrations of selected compounds, originally prepared at 1.0×10^{-2} M in 100% DMSO, were titered out to provide at least 12 individual concentrations. Twenty five (25) μ L of each solution (15% DMSO maximum) were added to wells in a 96-well plate, followed by 100 microliters (μ L) of a cocktail containing 100 nanograms (ng) of target protein at pH 7.0. Finally, 25 μ L of substrate solution was added and the plate (Immulon 2, Dynex) was read in 15 second intervals at 405 nanometers (nm) on a Spectramax 250 plate reader. IC₅₀ profiles and values were generated using the program Softmax.

Ligand A was shown to inhibit this enzyme with an IC₅₀ value of approximately 9.0×10^{-5} M. Subsequently, a similarity search resulted in the testing of about 10 structurally related compounds for enzyme inhibition. As shown in Figure 6, four of these compounds had IC₅₀ values between 2.0×10^{-5} M and 1.0×10^{-6} M. These very low affinity compounds can serve as lead chemical templates for the design of drugs directed against this molecular target.

Two-dimensional ¹H-¹⁵N HSQC spectra were subsequently recorded on [¹⁵N]target protein with and without Ligand A present to further investigate the interaction of this ligand with the protein. Chemical shift changes observed in the presence of Ligand A are delineated in Figure 7, which shows an overlay of the ¹H-¹⁵N HSQC spectra of protein alone and in the presence of a 10-fold

excess of ligand. Residues with the largest ligand-induced chemical shift changes are boxed.

In this study, a ligand that binds to an antibacterial target protein with a dissociation constant of less than about 2.0×10^{-4} M was identified from a small library of compounds. No prior knowledge of what types of ligands ought to bind to this protein was used. The identified ligand was shown to inhibit this enzyme with an IC_{50} value of approximately 9.0×10^{-5} M. Subsequently, a similarity search based on the structure of this NMR-identified ligand resulted in the testing of about 10 structurally related compounds for enzyme inhibition. Four of these compounds had IC_{50} values between about 2.0×10^{-5} M and about 1.0×10^{-6} M. These very low affinity compounds can serve as lead chemical templates for the design of drugs directed against this molecular target. More extensive NMR experiments, using isotopically-enriched target protein, concluded that the compounds identified as lead chemical templates do in fact bind to the active site of the target protein.

Example 3. Use of NMR Spectroscopy to Identify a Lead Chemical Template for an Antiviral Target Protein

Numerous protein targets are amenable to this NMR process of identifying a lead chemical template. In this example, the technique is illustrated for an antiviral target protein with a monomer molecular weight of approximately 8 kDa that exists as a dimer in solution. This target protein was screened using an NMR screening library and flow NMR spectroscopy.

All solutions containing the antiviral target protein were buffered with 2.0×10^{-2} M phosphate at pH 6.5. The protein used for the 1D screening and dissociation constant determination experiments was unlabeled, while that used for the 2D 1H - ^{15}N HSQC experiments was ^{15}N -enriched.

One-dimensional relaxation-edited 1H NMR spectra of samples containing a mixture of the target protein and a given compound combination were recorded in 2H_2O solution on a Bruker AMX-400 spectrometer. The spectrometer was equipped with a shielded magnet, a Gilson sample handler, and a 5 mm (250 μL sample cell) flow-injection NMR probe. A spin lock time of 350 milliseconds was used. The screening experiments were carried out on

solutions that were 3.8×10^{-5} M target protein and 5.0×10^{-5} M of each ligand. All solutions were contained in a 96-well plate and were delivered to the 5 mm flow-injection probe using the Gilson sample handler. The library used for the screening process was expanded from that described in the first two examples. It 5 contained approximately 300 compounds grouped into 32 separate mixtures.

Two-dimensional ^1H - ^{15}N HSQC spectra were recorded in $^1\text{H}_2\text{O}$ solution on a Bruker DRX-500 spectrometer. Samples contained 8.3×10^{-4} M target protein alone or in the presence of a given ligand.

Ligand dissociation constants were estimated by determining relative 10 diffusion coefficients for target protein alone, ligand in the absence of target protein, and ligand in the presence of target protein (Lennon et al., *Biophys. J.*, **67**, 2096-2109 (1994)). Relative diffusion coefficients were determined using pulsed-field-gradient NMR experiments incorporating a bipolar longitudinal eddy-current delay sequence (Wu, *J. Magn. Reson. Ser. A*, **115**, 260-264 (1995)).

15 One-dimensional relaxation-edited ^1H NMR spectroscopy was used to screen the expanded small molecule library for binding to this antiviral target protein in a manner analogous to that previously described in the first two examples. With this technique, a reduction in resonance intensity is observed if a compound interacts with the target protein, thus identifying it as a ligand. For 20 most of the compound combinations in the presence of the antiviral target protein, there was little or no reduction in resonance intensity with the 350-millisecond spin-lock time. However, for some of the compound combinations, the intensities of resonances corresponding to one of the compounds in the mixture were significantly reduced. The results from one such compound 25 combination are described here.

As a control, the 1D relaxation-edited ^1H NMR spectrum of a certain mixture in the absence of protein is shown in Figure 8A. All resonances are observed with full intensity. The corresponding 1D relaxation-edited ^1H NMR spectrum acquired in the presence of the antiviral target protein is shown in 30 Figure 8B. The intensities of all resonances corresponding to a single compound in Figure 8B are clearly reduced in the presence of the antiviral target protein. This indicates that this compound is binding to the protein. The binding is

specific to the antiviral target protein since the resonance intensities are not reduced in the presence of other protein targets that have been screened.

In a separate set of experiments that use pulsed-field-gradient techniques to determine relative diffusion coefficients, the dissociation constant for the 5 identified ligand was estimated by NMR measurements to be approximately 40 μM.

Two-dimensional ^1H - ^{15}N HSQC spectra were subsequently recorded on [^{15}N]target protein with and without the identified ligand present to further investigate the interaction of this ligand with the protein. Chemical shift changes 10 observed in the presence of this ligand are delineated in Figure 9, which shows an overlay of the ^1H - ^{15}N HSQC spectra of protein alone and in the presence of ligand. Residues with the largest ligand-induced chemical shift changes are labeled.

15 Example 4. Screening of Compound Libraries for Protein Binding Using
Flow-Injection NMR Spectroscopy

Introduction

Flow NMR spectroscopy techniques are becoming increasingly utilized in drug discovery and development (B. J. Stockman, *Curr. Opin. Drug Disc. Dev.*, **3**, 269-274 (2000)). The technique was first applied to couple the separation characteristics of liquid chromatography with the analytical capabilities of NMR spectroscopy (N. Watanabe et al., *Proc. Jpn. Acad. Ser B*, **54**, 194 (1978)). Since then, HPLC-NMR, or LC-NMR as it is more commonly referred to, has been broadly applied to natural products biochemistry, drug 20 metabolism and drug toxicology studies (J. C. Lindon et al., *Prog. NMR Spectr.*, **29**, 1 (1996); J. C. Lindon et al., *Drug. Met. Rev.*, **29**, 705 (1997); B. Vogler et al., *J. Nat. Prod.*, **61**, 175 (1998); and J.-L. Wolfender et al., *Curr. Org. Chem.* **2**, 575 (1998)). The wealth and complexity of data made available from the latter two applications have created the potential for NMR-based metabolomics to 25 complement genomics and proteomics (J. K. Nicholson et al., *Xenobiotica*, **29**, 1181 (1999)). Stopped-flow analysis in LC-NMR, where the chromatographic flow is halted to obtain an NMR spectrum with higher signal-to-noise and then restarted when the spectrum has finished collecting, was the forerunner to the 30

flow-injection systems that will be described here. The largest difference between the two systems is that one includes a separation component (LC column) and the other does not. The rapid throughput possible for combinatorial chemistry samples and protein/small molecule mixtures has allowed flow-

5 injection NMR methods to impact medicinal chemistry and protein screening (P. A. Keifer, *Drugs Fut.*, **23**, 301 (1998); P. A. Keifer, *Drug Disc. Today*, **2**, 468 (1997); P. A. Keifer, *Curr. Opin. Biotech.*, **10**, 34 (1999); K. A. Farley et al., SMASH'99, Argonne, IL, 15-18 August 1999; and A. Ross et al., *Biomol. NMR*, **16**, 139 (2000)).

10 Changes in chemical shifts, relaxation properties or diffusion coefficients that occur upon the interaction between a protein and a small molecule have been documented for many years (for recent reviews see M. J. Shapiro et al., *Curr. Opin. Drug. Disc. Dev.*, **2**, 396 (1999); J. M. Moore, *Biopolymers*, **51**, 221 (1999); and B. J. Stockman, *Prog. NMR Spectr.*, **33**, 109 (1998)). Observables

15 typically used to detect or monitor the interactions are chemical shift changes for the ligand or isotopically-enriched protein resonances (J. Wang et al., *Biochemistry*, **31**, 921 (1992)), or line broadening (D. L. Rabenstein, et al., *J. Magn. Reson.*, **34**, 669 (1979); and T. Scherf et al., *Biophys. J.*, **64**, 754 (1993)), change in sign of the NOE from positive to negative (P. Balaram et al., *J. Am. Chem. Soc.*, **94**, 4017 (1972); and A. A. Bothner-By et al., *Ann. NY Acad. Sci.* **222**, 668 (1972)), or restricted diffusion (A. J. Lennon et al., *Biophys. J.* **67**, 2096 (1994)) for the ligand. For the most part, these studies have focussed on protein/ligand systems where the small molecule was already known to be a ligand or was assumed to be one. In the last several years, however, the work of

20 the Fesik (S. B. Shuker et al., *Science*, **274**, 1531 (1996); and P. J. Hajduk et al., *J. Am. Chem. Soc.*, **119**, 12257 (1997)), Meyer (B. Meyer et al., *Eur. J. Biochem.*, **246**, 705 (1997)), Moore (J. Fejzo et al., *Chem. Biol.*, **6**, 755 (1999)), Shapiro (M. Lin et al., *J. Org. Chem.*, **62**, 8930 (1997)), and Dalvit (C. Dalvit et al., *J. Biomol NMR*, **18**, 65-68 (2000)) labs has demonstrated the applicability of

25 these same general methods as a screening tool to identify ligands from mixtures of small molecules.

These screening protocols typically involve the preparation of a series of individual samples in glass NMR tubes and the use of an autosampler to achieve

reasonable throughput. Variations in volume or positioning that occur during sample preparation or tube insertion can necessitate tuning and calibration of the probe between each sample, thereby reducing throughput of data collection.

By contrast, flow-injection NMR has several advantages. The stationary 5 flow cell provides uniform locking and shimming from one sample to the next, and, with the radio frequency coils mounted directly onto the flow cell's glass surface, high sensitivity. Fast throughput of data collection is thus possible. Use of a liquid handler to prepare and inject samples, such as the Gilson 215 liquid handler used on Bruker and Varian systems, allows the potential for on-the-fly 10 sample preparation (A. Ross et al., *J. Biomol. NMR*, **16**, 139 (2000)), thus maximizing sample integrity and uniformity. Since the use and/or re-use of glass NMR tubes is avoided, costs are minimized.

Data Acquisition Hardware and Software

15 A typical Flow NMR system consists of a magnet, an NMR console, a computer workstation, a Gilson sample handler, and a flow-injection probe. Two vendors currently offer complete flow-injection systems: Bruker Instruments and Varian Instruments. In addition, the Nalorac Corporation manufactures an LC probe that can also be used for flow-injection NMR 20 screening. A schematic of the Bruker Efficient Transport System (BEST) manufactured by Bruker Instruments is shown in Figure 10. The Gilson 215 sample handler supplied by Bruker is equipped with two Rheodyne 819 valves. The first valve is attached to a 5 ml syringe, the needle capillary in the sample handler injection arm, the bridge capillary, the waste reservoir, and the second 25 valve. The second Rheodyne valve is attached to the input and output of the probe, the source of nitrogen gas, the first valve, and the injection port. FEP Teflon tubing is used in each of the connections with the exception of the gas connection, which uses PEEK tubing.

A sample is injected into the Bruker probe by filling the needle capillary 30 and transferring the sample into the inlet tubing for the probe using the second Rheodyne valve. In quick mode, the next sample is loaded into the tubing during the spectral acquisition of the previous sample. When the spectral acquisition has completed, the first sample exits the probe through the outlet capillary. This

TOP SECRET - 5724100

action pulls the next sample into the probe through the inlet port and spectral acquisition can immediately begin. Quick mode acquisition can save approximately one minute per sample from the time it would take to load each sample individually. However, sample recovery is not currently an option with 5 this method. In order to recover a sample, each sample is injected individually using normal mode acquisition. The sample is recovered by selecting either nitrogen gas or the syringe to pull the sample back from the probe through the inlet tube. The sample can then be returned to the Gilson liquid handler into its original well or into a new 96 well plate. A recovery unit has recently been 10 added to the BEST system to improve the efficiency of recovery of the syringe by using the nitrogen gas to create a back pressure on the sample.

Two useful accessories available for the BEST system are a Valvemate solvent switcher and a heated transfer line. The solvent switcher was added to the flow system for the combinatorial chemist who may want to analyze samples 15 in various organic solvents, but it can also be used for a library screen to vary buffer conditions or to clean the probe out with an acid or a base. The heated transfer line is used to equilibrate the sample temperature to the probe temperature during sample transfer. Both the inlet and output capillary transfer lines are threaded through the heated transfer line. This feature is desirable when 20 the spectral analysis time is short and a high throughput of samples is required. In the ideal case, data acquisition using this accessory can begin immediately after the sample enters the probe. Some samples may still require a temperature equilibration period after entering the probe.

The setup of the Versatile Automated Sample Transport (VAST) system 25 produced by Varian is similar to the Bruker system. The VAST system consists of a Gilson 215 liquid handler, a Varian NMR flow probe, an NMR console, and a Sun workstation. The Gilson liquid handler supplied by Varian is equipped with a single Rheodyne 819 valve and is connected to the NMR flow probe with 0.010 inch inside diameter PEEK tubing (P. A. Keifer et al., *J. Comb. Chem.*, 2, 30 151 (2000)). In the Varian system design, the sample handler injects a specified volume of sample into the probe, the data is acquired, and then the flow of liquid through the tubing is reversed and the sample is returned to its original vial or well. The return of the sample to the Gilson by the syringe pump is assisted by a

Valco valve and nitrogen gas which supply some backpressure on the outlet portion of the Varian flow probe. With the VAST system setup, the probe is rinsed just prior to sample injection and then is dried with nitrogen gas to minimize dilution of the sample during injection. The Varian design gives

5 excellent sample recovery without dilution, but it is strongly recommended that samples be filtered to prevent clogging of the capillary transfer lines (P. A. Keifer et al., *J. Comb. Chem.*, **2**, 151 (2000)).

Flow NMR systems are ideally suited for use with the shielded magnets manufactured by Bruker Instruments or Oxford Magnets. Actively shielding a

10 600 MHz magnet reduces the radial 5 gauss line from approximately 4 meters to less than 2 meters, which allows the Gilson liquid handler to be placed significantly closer to the magnet. This reduces the length of tubing needed between the Rheodyne valve and the flow-injection probe and minimizes the sample transfer time. The potential for clogging and sample dilution are

15 concomitantly reduced.

Bruker uses two software packages to run the BEST system: BEST Administrator and ICONNMR (Bruker Instruments, AMIX, BEST and ICONNMR software packages). The BEST administrator is activated by typing the command 'BESTADM' in XWINNMR. This portion of the software is used

20 during method generation and optimization. Samples are injected into the probe one at a time and data is collected under XWINNMR. Early versions of the BEST software utilized three separate programs: CFBEST, SUBEST, and OTBEST. These functions were recently combined under the single software package, BEST Administrator. In addition, the parameters available for

25 customization have been greatly expanded to include automated solvent switching and method switching, which were not available in earlier versions of the software. The software package ICONNMR is used after a flow method has been optimized with the BEST administrator. This package is setup for full automation and is the same software used with automated NMR tube sample

30 changers. In a similar fashion, Varian software uses the command 'Gilson' to generate a method before sample injection and data acquisition is initiated using Enter/Autogo in VNMR (Varian NMR Systems, VNMR software package).

TOP SECRET - FEDERAL BUREAU OF INVESTIGATION

Flow Probe Calibration and System Optimization

In addition to the normal 90° pulse lengths and power levels which are calibrated for any NMR probe, several additional calibrations are required for a flow probe. The three additional volumes required to calibrate a Bruker flow probe are shown schematically in Figure 11 (Bruker Instruments, AMIX, BEST and ICONNMR software packages). The first volume calibrated is the total probe volume. This can be accomplished by injecting a colored liquid into the inlet of a dry probe with a syringe and watching for the liquid to appear in the outlet port (approximately 700-800 µL for a 5 mm flow probe). With the Varian system, the system filling volume also includes the capillary tubing that connects the injector port to the flow probe (P. A. Keifer et al., *J. Comb. Chem.*, **2**, 151 (2000)). This volume is used to calculate the distance required to reposition a sample from the Gilson sample handler to the center of the flow cell in the probe.

The second volume calibrated is the flow cell volume. This is the volume of liquid required to fully fill the coil around the flow cell. The three flow probe vendors (Bruker, Varian, and Nalorac) have probes available with active volumes ranging from 30-250 µL. The stated volume of the flow cell in a 5 mm Bruker flow probe is 250 µL, but it was calibrated to be approximately 300 µL. This volume can be calibrated by making repeated injections of a standard sample, starting with a volume less than the stated active volume of the probe, and collecting a 1D ¹H NMR spectrum. The injection volume can then be increased incrementally until no further improvement in signal-to-noise is observed.

In addition to the two probe volume calibrations already discussed, Bruker software also includes a third volume for calibration. This volume, referred to as the positioning volume, is used to optimize the centering of a sample in the flow cell. Early versions of ICONNMR software (prior to 3.0.a.9) did not include the ability to set the positioning volume. Rather, Bruker literature suggested that the flow cell volume should be roughly doubled to insure that the sample would completely fill the coil (Bruker Instruments, AMIX, BEST and ICONNMR software packages). Fortunately, this is no longer necessary. The positioning volume can now be used to optimize the sample

position. This calibration reduced the sample size required for injection from 450 µL in the first few protein screens to 300 µL for current screens using a Bruker 5 mm flow probe with an active volume of 250 µL. Optimization of this parameter minimized the sample volume required for each spectrum.

5 Importantly, this significantly reduced the total amount of protein (or other target) at a given concentration needed to screen our small molecule library. The positioning volume can be optimized by collecting a series of spectra on a standard sample. In each spectrum collected, the positioning volume can first be varied by large increments (50-100 µL) to get a rough estimate of the volume.

10 An example of three such spectra is shown in Figure 12. The positioning volume can then be varied in smaller increments (10-25 µL) to identify the best volume for this parameter. The best signal-to-noise was obtained for our 5 mm Bruker flow probe on a DRX-600 when the positioning volume was set to +25 µL, but this volume is probe specific and is calibrated for each flow probe.

15 The optimization of a flow-injection system for screening has three main objectives. The first objective is to transfer an aqueous sample to the center of the flow cell for analysis using the parameters determined during the flow probe calibration described above. The second objective is to reposition a sample from the Gilson liquid handler into the flow-injection probe without bubbles and with minimal sample dilution. This can be achieved by using nitrogen as a transfer gas (which keeps the system under pressure) and by using a series of leading and trailing solvents. In our experiments, we typically use 150 µL of $^2\text{H}_2\text{O}$ as a leading solvent, 20 µL of nitrogen gas, 300 µL of sample, 20 µL of nitrogen gas, and 100 µL of $^2\text{H}_2\text{O}$ as a trailing solvent. Alternatively, a larger volume of

20 sample can be used in place of the push solvents. The third objective is to determine a cleaning procedure which would reduce sample carry-over to less than 0.1%. Typically, this involves rinsing the probe with a predetermined volume of water. The rinse cycle can also be followed by a dry cycle, in which the capillary lines and flow probe are dried with nitrogen gas to further minimize

25 sample dilution. In our experiments, we typically use a 1-mL wash volume followed by a 30 second drying time with nitrogen gas.

PROTETED BY PATENT

Design of Small Molecule Screening Libraries

With the increasing prevalence of extremely high throughput screening equipment in the pharmaceutical industry, it may seem counter intuitive to suggest screening smaller collections of compounds in an NMR-based assay.

5 However, a correlation between the quality of hits obtained and the number of compounds screened has not been well documented. In fact, compounds are typically added to screening collections not to simply increase their numbers, but to increase the diversity and quality of the compound collection. Thus, if one could find suitable hits from a smaller collection of well-chosen compounds, it

10 may not be necessary to expend the time and chemical resources to screen the entire compound library against every single target. Hits so identified could then be used to focus further screening efforts or to direct combinatorial syntheses, thus saving both time and chemical resources, as shown schematically in Figure 1. An NMR-based screen, like other binding assays, has the advantage in that a

15 high throughput functional assay does not need to be developed. This will become increasingly important as more and more targets of interest to pharmaceutical research are derived from genomics efforts and thus may not have a known function that can be assayed.

Several types of libraries are possible: broad screening libraries applicable to many types of target proteins, directed libraries that are designed with the common features of an active site in mind that might be useful for screening a series of targets from the same protein class, such as protease enzymes, and “functional genomics” libraries composed of known substrates, cofactors and inhibitors for a diverse array of enzymes that might be useful for defining the function of genomics-identified targets.

Ideally, the size and content of a broad screening library should be such that screening can be accomplished in a day or two with a favorable chance of identifying several hits for each of the target proteins to be screened. Rather than just randomly choosing a subset library, several rationale approaches have been implemented. These include the SHAPES library developed by Fejzo and coworkers that is composed largely of molecules that represent frameworks commonly found in known drug molecules (J. Fejzo et al., *Chem. Biol.*, 6, 755 (1999)), drug-like or lead-like libraries, and diversity-based libraries. A number

of studies have recently appeared that discuss the properties of known drugs and methods to distinguish between drug-like and non-druglike compounds (G. W.

Bemis et al., *J. Med. Chem.*, **39**, 2887 (1996); C. A. Lipinski et al., *Adv. Drug Del. Rev.*, **23**, 3 (1997); Ajay et al., *J. Med. Chem.*, **41**, 3314 (1998); J. Sadowski et al., *J. Med. Chem.*, **41**, 3325 (1998); A. K. Ghose et al., *J. Comb. Chem.*, **1**, 55 (1999); J. Wang et al., *J. Comb. Chem.*, **1**, 524 (1999); and G. W. Bemis et al., *J.*

5 et al., *J. Med. Chem.*, **41**, 3325 (1998); A. K. Ghose et al., *J. Comb. Chem.*, **1**, 55
10 (1999); J. Wang et al., *J. Comb. Chem.*, **1**, 524 (1999); and G. W. Bemis et al., *J.
Med. Chem.*, **42**, 5095 (1999)). Superimposing drug-like (E. J. Martin et al., *J.
Comb. Chem.*, **1**, 32 (1999)) or lead-like (S. J. Teague et al., *Angew. Chem. Int.
Ed.*, **38**, 3743 (1999)) properties on a diversity-selected compound set may yield
the best library of compounds. The distinction of lead-like is important since the
NMR-based assay is designed to identify weak-affinity compounds that will
most likely gain molecular weight and lipophilicity to become drug candidates or
even lead chemical templates (S. J. Teague et al., *Angew. Chem. Int. Ed.*, **38**,
3743 (1999)).

15 Development and expansion of our lead-like NMR screening library to mimic the structural diversity of our larger compound collection has made use of the DiverseSolutions software for chemical diversity (R. S. Pearlman et al., *Persp. Drug Disc. Des.*, **9/10/11**, 339 (1998)). In this approach, each compound is described by a set of descriptors, which are metrics of chemistry space. Six
20 orthogonal descriptors, related to substructures as opposed to the entire molecule, are often used. While the descriptors to use can be automatically chosen to maximize diversity, typically there are two each corresponding to charge, polarizability and hydrogen-bonding. A cell-based diversity algorithm is employed to divide the descriptor axes into bins and thus into a lattice of
25 multidimensional hypercubes. As an example of how this can be used to construct or expand a small screening library, consider the selection of 1,000 compounds from a compound library of 250,000 compounds. First, the cell-based algorithm is used to partition the 250,000 compounds into approximately 1,000 cells. The number of compounds per cell will vary and some will be
30 empty. Maximum structural diversity will be obtained by taking one compound from each occupied cell (and as close to the center as possible). The actual compounds chosen are based on desirable lead-like properties such as low molecular weight and hydrophilicity as well as availability and chemical non-

reactivity as explained below. Diversity voids, as exemplified by empty cells, can be filled from external sources or by chemical syntheses if desired. Identifying and filling diversity voids is important since larger compound collections are often heavily weighted in certain classes of compounds stemming 5 from earlier research projects.

An example of diversity-based subset selection using these methods is shown in Figure 13. Here, the 6,436 compounds from the Comprehensive Medicinal Chemistry index have been divided into 2,012 cells to maximize diversity using five chemistry-space descriptors. The two-dimensional 10 representation projected onto the hydrogen bond acceptor and charge BCUT axes is shown in gray. The black squares correspond to the 1,474 lead-like compounds (molecular weight less than 350 and $1 < c\text{LogP} < 3$) contained in the CMC index. A total of 806 of the 2,012 cells were occupied by lead-like 15 compounds. A similar approach could be used to select diverse, lead-like compounds from a large corporate compound collection.

The cell concept of structural space is quite useful after the screening is complete. When a hit is identified, other compounds from the same or nearby cells are obvious candidates for secondary assays. One can think of this as the gold mine analogy: when gold is struck, the search is best continued in close 20 proximity.

In addition to structural diversity, there are other characteristics that can be considered when selecting the subset molecules. These include purity, identity, reactivity, toxicological properties, molecular weight, water solubility, and suitability for chemical elaboration by traditional or combinatorial methods. 25 It makes sense to populate the screening library with compounds of high integrity that are not destined for failure down the road. Time spent upfront to insure purity and identity with LC-MS or LC-NMR analyses will save resources downstream. Filtering tools can be used to avoid compounds that are known to be highly reactive, toxic, or to have poor metabolic properties. Lack of reactivity 30 is important since compounds can be screened more efficiently as mixtures.

Like other labs (S. B. Shuker et al., *Science*, **274**, 1531 (1996); B. Meyer et al., *Eur. J. Biochem.*, **246**, 705 (1997); J. Fejzo et al., *Chem. Biol.*, **6**, 755 (1999); and M. Lin et al., *J. Org. Chem.*, **62**, 8930 (1997)) we typically pool our selected

100-112-113-114-115-116-117-118-119-120

small molecules into mixtures of 6-10 compounds for screening (K. A. Farley et al., SMASH'99, Argonne, IL, 15-18 August 1999).

Compounds chosen for our diversity library are lead-like as opposed to drug-like. It is often the case that chemical elaborations to improve affinity also increase molecular weight and decrease solubility (S. J. Teague et al., *Angew. Chem. Int. Ed.*, **38**, 3743 (1999)). The molecular weight of the compounds therefore should preferably not exceed about 350. Since most hits obtained will have affinities for their target in the approximately 100 μM range, low molecular weight will leave room for chemical elaboration to build in more affinity and selectivity. Using larger molecular weight drug-like compounds would not substantially improve affinity of the hits and could easily preclude obtaining lead chemical templates of reasonable size. Lead-like hits that are reasonably water soluble allow for chemical elaboration that results in modest increased lipophilicity of the final therapeutic entity (S. J. Teague et al., *Angew. Chem. Int. Ed.*, **38**, 3743 (1999)). Water solubility is also important since it enhances the potential success of downstream studies such as calorimetry, enzymology, co-crystallization and NMR structural studies. Compound solubility is especially important for flow-injection NMR methods in order to prevent clogging of the capillary lines.

Compounds should also be chosen with their suitability for chemical elaboration by traditional or combinatorial chemistry methods in mind. Hits with facile handles for synthetic chemistry will be of more interest and will allow more efficient use of often limited medicinal chemistry resources.

25 Relaxation-Edited or WaterLOGSY-Based Flow-Injection NMR Screening Methods

Calibration and validation of the flow system and creation of a small-molecule screening library yields an automated system that is ready to screen new targets. A protein target can be analyzed for protein-ligand interactions using relaxation-editing methods by adding sufficient protein to each well of the 96-well library plate to give a 1:1 (protein:ligand) ratio at a concentration of approximately 50 μM . Homogeneous sample dispersion throughout the well can

be facilitated by agitating the plate on a flat bed shaker. Screening at this concentration allows a decent 1D ^1H NMR spectrum to be acquired in about 10 minutes. In our experience, this concentration of target and small molecule requires identified ligands to have affinities on the order of approximately 200

5 μM or tighter.

Once the screening plate has been prepared, the Gilson liquid sample handler transfers samples from 96-well plates into the flow-injection probe and if desired, returns the samples back into either the original 96-well plate or a new plate. Once the sample is in the magnet, spectra that can detect changes in

10 chemical shifts, relaxation properties, or diffusion properties can be collected. In our relaxation-edited NMR screening assay, two 1D relaxation-edited ^1H NMR spectra are collected: one spectrum is collected on the ligand mixture in the presence of protein and the second, control spectrum is collected on the ligand mixture in the absence of protein. Ligands are identified as binding to a target

15 when their resonances are greatly reduced when compared to a relaxation-edited spectrum collected in the absence of protein as illustrated in Figure 14. In this example, the target protein was a genomics-derived protein of unknown function.

Ligand binding can be confirmed by collecting a 1D relaxation-edited ^1H

20 NMR spectrum of each individual ligand that was identified as binding to the protein in a given mixture as shown in Figure 15. In addition, the binding constant of the protein/ligand interaction can be estimated using 1D diffusion-edited spectra of the ligand in the presence and absence of protein (A. J. Lennon et al., *Biophys. J.*, **67**, 2096 (1994)). If labeled protein is available, a 2D ^1H - ^{15}N

25 HSQC spectrum can also be obtained to locate the ligand binding site on the protein (J. Wang et al., *Biochemistry*, **31**, 921 (1992); and S. B. Shuker et al, *Science*, **274**, 1531 (1996)). In cases where the protein is small enough and structural characterization of the binding interaction is desired, further experiments can be carried out using ^{15}N and/or ^{13}C / ^{15}N protein/ligand

30 complexes.

When binding is detected using the WaterLOGSY technique, sample preparation and use of the flow-injection apparatus is identical, except that

TOP SECRET - FEDERAL BUREAU OF INVESTIGATION

extremely low levels of target are used (1-10 μ M) with ratios of ligand to target of 100:1 to 10:1. For data analysis, binding compounds are distinguished from nonbinders by the opposite sign of their water-ligand NOEs. In contrast to the relaxation-edited technique, only a single WaterLOGSY spectrum is used for 5 each ligand mixture. There is no need to collect a reference spectrum in the absence of target protein. An example is illustrated in Figure 16 for a mixture of compounds and a different protein. In the WaterLOGSY spectrum shown in Figure 16, binding compounds have resonances of opposite intensity (sharp positive peaks) than nonbinders (near zero intensity or sharp negative peaks).
10 Residual protein resonances are also of positive intensity.

Data Analysis

The development of flow probes has facilitated the transition to high-throughput NMR and has made possible the routine collection of tremendous 15 volumes of data. Recent software developments have advanced the automated handling of large data sets collected on combinatorial chemistry libraries (P. A. Keifer et al., *J. Comb. Chem.*, **2**, 151 (2000); Bruker Instruments, AMIX, BEST and ICONNMR software packages; Varian NMR Systems, VNMR software package; and Williams A, Book of Abstracts, 218th ACS National Meeting 20 (1999)). Visualization of results in a 96-well format allows rapid evaluation of the data sets. The integration of features such as this into a software package tailored more for data reduction and evaluation of library screening data sets parallels the combinatorial chemistry software development but remains slightly behind. However, recent advancements that have been made for combinatorial 25 chemistry data analyses portend similar developments for the automation of protein binding screening data.

In our 1D relaxation-edited ^1H NMR data sets, one can simply identify the ligand resonances by inspection since their intensity is reduced in the presence of protein as shown in Figure 14. In our WaterLOGSY data sets, 30 binding compounds are distinguished from nonbinders by the opposite sign of their water-ligand NOEs as observed in Figure 15. In either case, comparison to an assigned small molecule control spectrum are made to identify the compound associated with the indicated resonances.

Other labs have relied on difference spectra to analyze relaxation- or diffusion-edited 1D ^1H NMR data sets (P. J. Hajduk et al., *J. Am. Chem. Soc.*, **119**, 12257 (1997); N. Gonnella et al., *J. Magn. Reson.*, **131**, 336 (1998); and A. Chen et al., *J. Am. Chem. Soc.*, **122**, 414 (2000)). After a series of spectral 5 subtractions, the resulting spectrum represents the resonances of the compounds that bind to the protein. Two factors that pose problems are line broadening and shifting resonances, both of which can lead to subtraction artifacts. Changes in intensity can also add the need for a scaling factor in the data analysis step. These additional steps, which can vary from one spectra to the next, make 10 strategies for automated data analysis complex.

Data analysis for 2D screening methods typically involves either the analysis of protein chemical shift perturbations indicative of ligand binding (A. Ross et al., *J. Biomol. NMR*, **16**, 139 (2000); and S. B. Shuker et al., *Science*, **274**, 1531 (1996)), or the analysis of changes in signals from the small molecules in 15 NOE or DECODES spectra indicative of binding (B. Meyer et al., *Eur. J. Biochem.*, **246**, 705 (1997); J. Fejzo et al., *Chem. Biol.*, **6**, 755 (1999); and M. Lin et al., *J. Am. Chem. Soc.*, **119**, 5249 (1997)). While a series of 2D ^1H - ^{15}N HSQC spectra can be compared manually, automated analysis using both non-statistical and statistical approaches of a series of ^1H - ^{15}N HSQC spectra acquired 20 with flow-injection NMR methods was recently demonstrated (A. Ross et al., *J. Biomol. NMR*, **16**, 139 (2000)). AMIX was used for the non-statistical analysis by comparing spectra collected in the presence of single compounds to the reference spectrum of the protein alone. Then, using bucketing calculations for data reduction, a table ranked by the correlation coefficient was generated. No 25 correlations were observed using the bucketing calculations alone.

Subsequently, integration patterns for all 300 small molecule spectra were analyzed by AMIX to generate a data matrix of N integration regions times 300. A statistical software package, UNSCRAMBLER 6.0, was then used to analyze this data matrix using principal components analysis. Two classes of spectral 30 changes were observed. Ultimately, one class was found to correspond to pH changes caused by certain small molecules while the other class corresponded to small molecules binding to the target protein (A. Ross et al., *J. Biomol. NMR*, **16**, 139 (2000)).

Data reduction is an important aspect for handling the amounts of data generated if high-throughput screening by NMR is to be successful. Non-statistical methods such as the bucketing calculations of AMIX (Bruker Instruments, AMIX, BEST and ICONNMR software packages) or the database 5 comparisons of ACD (Williams A, Book of Abstracts, 218th ACS National Meeting (1999)) compare chemical shift, multiplicity, integration regions and patterns to give correlation factors between spectra. These software packages can be used for data reduction of both one- and two-dimensional data. Prediction software is also available to help aid in interpretation of data sets. Statistical 10 methods such as principal components analysis can be used to analyze data for other correlations that are not apparent using non-statistical methods alone. In the case of 2D ^1H - ^{15}N HSQC data, an adaptive, multivariate method that incorporates a weighted mapping of perturbations to correlate information within a spectrum or across many spectra has also been described (F. Delaglio, CHI 15 Conference on NMR Technologies: Development and Applications for Drug Discovery, Baltimore, MD, 4-5 November 1999).

Comparison of Flow vs. Traditional Methods

The advantage of working with samples in the flow NMR screening 20 environment is that each set of spectra are collected on samples that are at the same concentration. This accelerates spectral acquisition considerably. Since the samples are fairly homogenous, many of the routine tasks need to be completed on only the first sample: probe tuning, ^1H 90° pulse calibration, receiver gain, number of transients, locking, and gradient shimming. On 25 subsequent samples, these steps can be omitted, although simplex shimming of Z_1 and Z_2 can still be used with multi-day acquisitions.

Prerequisites for a high-throughput assay include rapid data collection, sample-to-sample integrity and minimal costs. Flow NMR techniques have been developed with each in mind. For 1D ^1H NMR screening experiments, the 30 process of removing the previous sample from the flow cell, rinsing the flow cell, injecting the next sample, allowing for thermal equilibration, automating solvent suppression and acquiring the data can take less than 10 minutes. In practice, the use of this procedure is two to three times faster than a sample

changer with conventional NMR tubes. If compounds were screened in mixtures of 10, this results in a throughput of about 1,500 compounds per day. Use of a liquid handler, such as the Gilson 215 typically employed by Bruker and Varian flow NMR systems, can simplify the preparation of samples as well. Ross and 5 coworkers have demonstrated on-the-fly sample preparation by using the liquid handler to mix the protein to be screened with the small molecule immediately prior to injection (A. Ross et al., *J. Biomol. NMR*, **16**, 139 (2000)). Sample conditions can thus be highly standardized with the resulting spectra very consistent and reproducible. Even if target protein is added manually to pre- 10 plated screening libraries, the amount of pipetting is still less than if using NMR tubes. Recurring expenses associated with purchasing and/or cleaning NMR tubes are eliminated with flow-injection NMR methods. The cost of the 96-well microtitre plates is insignificant compared to NMR tubes.

In other embodiments, the methodologies described above also can be 15 used to determine the potential biological roles of proteins having previously unknown function. In today's era of high throughput genome sequencing, complete genomes of tens of organisms have already been sequenced and work on hundreds more is in progress. This has led to identification of thousands of new proteins. The potential of these proteins to act as drug targets cannot be fully 20 assessed without the knowledge of the protein's function and importance in biological processes.

Historically, functional assays, such as those described above, have been used to identify compounds that bind to proteins having known function (drug targets), which eventually become drug candidates. The NMR binding assays 25 described above can be used to identify compounds that bind to proteins of unknown function. Identifying which types of compounds bind to a protein can help in understanding the previously unknown biological and/or biochemical function of the protein. Specific interactions between macromolecules and smaller molecular weight ligands are important in all biochemical processes. 30 Enzymes require specific binding of cofactors and/or substrates to carry out the reactions that they catalyze. Inhibitors are designed to specifically bind enzymes and receptors in or around the active site, and they often are analogous to substrates or cofactors.

Specific interactions are necessary for the proteins to carry out their functions. Hence, identifying which compounds bind to proteins of unknown function can provide clues about that protein's function. For example, a hypothetical protein function may be identified by characterizing those 5 compounds that bind to the protein in terms of their function as inhibitors, cofactors, or substrates of known proteins. NMR based binding assays can be used to identify which ligands in a screening library bind to the protein. Knowing what types of ligands bind to the protein helps to estimate the protein's function, which in turn, facilitates analyzing the protein's potential as a drug 10 target by creating a target priority list.

Screening Library Design

Several databases were searched to find known inhibitors, cofactors, and substrates of known proteins. Four hundred and thirty compounds were compiled through these searches. Small amounts (about 2-5 mgs.) of 220 15 compounds were obtained internally or from Sigma/Aldrich. All these compounds were tested for solubility and purity. The solubility tests involved assessment of the compounds to make a 50mM stock solution in either DMSO or 100mM phosphate buffer, pH 6.5. The solubility of the compounds was also tested at 100 μ M concentration in 100mM phosphate buffer pH 6.5, which is a 20 typical NMR binding assay condition. The purity of the compounds was checked by mass spectrometry and NMR spectroscopy.

The screening library finally contained 156 compounds, all of which passed the solubility and purity tests. These compounds had a range molecular weights from 46 to 1389 with average molecular weight being 301. These 25 compounds are also known to interact with a wide spread of enzyme classes covering a broad spectrum of metabolic pathways. Table 1 describes the distribution of the library compounds over the major enzyme classes. Of course, it is possible to add more compounds to this library as they are identified by their interactions with known proteins.

Enzyme class	Enzyme type	Number of compounds
1	Oxidoreductases	60
2	Transferases	34
3	Hydrolases	56
4	Lyases	32
5	Isomerases	13

Table 1. Distribution of compounds over major enzyme classes.

TOP SECRET - STAFF

5 Preparation of Mixtures

To improve screening efficiency, the library was compressed into 30 mixtures, each containing 4 to 7 compounds. The criteria for inclusion of compounds in the mixtures were non-reactivity with each other, and presence of 10 at least one unique resonance corresponding to each compound in the NMR spectrum of the mixture. The NMR spectra of each compound in the mixture were added together to create a theoretical spectrum of the mixture and it was compared with the actual NMR spectrum of the mixture. All theoretical and experimental spectra were consistent with each other indicating non-reactivity of 15 the compounds in the mixtures. There were two types of mixtures depending on compounds that were dissolved in DMSO or buffer to make stock solutions. The mixtures were prepared in 96 well plates, and stored at –80 °C till they were used for screening experiments.

20 Validation of library

Several proteins with known functions were used for validating the screening library. The proteins used for validation are listed in Table 2. The proteins were dissolved in 100mM phosphate buffer, pH 6.5 to make stock 25 solutions which were further diluted and mixed with the compound mixtures to make final concentration of 5-7 µM. The concentration of compounds was about

133 μ M in the final solution. The ratio of compound to protein concentrations was about 20:1.

Protein	Molecular weight (kDa)
γ -Chymotrypsin	22
Alcohol dehydrogenase	80
Carbonic anhydrase	29
Thrombin	34
Camphor Cytochrome P450	47
Transketolase	74
Lactate dehydrogenase	45

5 Table 2. Test proteins used in validation of the library.

NMR Screening experiments

NMR experiments for validating the functional genomics library with
10 proteins of known functions were conducted on a Bruker Avance 600 MHz
spectrometer equipped with 5mm FISEI flow probe and Gilson 215 liquid
sample handler. Binding was detected using the WaterLOGSY experiment.

Results from Thrombin screening experiments

15 The functional genomics screening library was screened against thrombin obtained from Sigma, which is one of the test proteins used for validation of the library. One assay mixture contained 133 μ M of N-alpha-dansyl-DL-tryptophan cyclohexylammonium salt (DPS) and 7 μ M of thrombin in 100 mM phosphate buffer, pH6.5. This mixture also contained Benzyl (S)-(-)-2-(1-pyrrolidinylcarbonyl)-1-pyrrolidinecarboxylate (ZPR), Chymostatin A (CSA), Tetrahydrofolic acid (C2F), Haloperidol (THK). Referring to Figure 17, the reference NMR spectrum of DPS is in the top panel while the WaterLOGSY spectrum of the mixture is shown in the bottom panel. The positive peak in the

WaterLOGSY spectrum indicates binding of DPS to thrombin. The peaks indicated by red asterisks in the WaterLOGSY spectrum correspond to peaks from the reference spectrum of DPS shown in the top panel.

The complete disclosures of the patents, patent documents, and

5 publications cited herein are incorporated by reference in their entirety as if each were individually incorporated. Various modifications and alterations to this invention will become apparent to those skilled in the art without departing from the scope and spirit of this invention. It should be understood that this invention is not intended to be unduly limited by the illustrative embodiments and

10 examples set forth herein. Such examples and embodiments are presented by way of example only with the scope of the invention intended to be limited only by the claims set forth herein as follows.

DOCKET # 6283.NCP2